

The complexity of chemical information: interview with Gary Wiggins

10



Gary Wiggins

Gary Wiggins was the head of the Indiana University (IU) Chemistry Library from 1976 to 2003. During the final four years of his professional career, he served as director of the Bioinformatics and Cheminformatics Programs in the IU School of Informatics, helping to create one of the first graduate programs in the United States that offer specialized training in cheminformatics. For many years, he taught courses in chemical information and science reference at IU. His textbook *Chemical Information Sources* was eventually converted to a Wikibook. Dr. Wiggins received several prestigious awards throughout his career, including the American Chemical Society Division of Chemical Information's Herman Skolnik Award and the Patterson-Crane Award of the ACS Columbus and Dayton Sections. He was also elected to the Special Libraries Association Hall of Fame. Much of his research involved the improvement of teaching information literacy to chemistry and science students and the improvement of communication among scientists.

Svetla Baykoucheva: *In May 1991, you started a chemical information discussion list in Indiana University. Through the years, this forum became an institution of its own, providing a medium for exchanging information and ideas and attracting people interested in chemical information, but who approached it from different perspectives. Looking back at the dynamics of the Chemical Information Sources Discussion List, what do you think was the impact of this unique forum on the evolution of chemical information as a discipline, and how did it benefit those who engaged in this discourse?*

Gary Wiggins: In this era of social media, it is surprising to me that an e-mail Listserv based on technology developed over 20 years ago is still thriving. In many ways, CHMINF-L is still *the* information source for everyone from chemists to

chemistry/science librarians and publishers. I think that the diversity of the audience was likely the single thing that made it so valuable. Chemists were able to learn of new information sources and the tricks of the trade from librarians. Librarians were able to find help from chemists and other colleagues when reference questions were too complicated for them to understand the underlying subtleties that kept them from finding the right answers. The publishers of scientific journals and databases could avail themselves of a free resource where they could assess information needs and do market research. In that sense, CHMINF-L served to identify areas where there were gaps in chemical information and knowledge and helped to focus publishers on tools that were needed by the chemical information community. The benefit to those who participated in the discourse can be judged by the continuing support of the existing subscribers (over 1300), quite a number of whom have been on the list since its inception.

SB: *The American Chemical Society (ACS) created the Division of Chemical Information (CINF) 70 years ago; this Division has been publishing a journal, the Chemical Information Bulletin, for 65 years; Indiana University has a graduate program in chemical information; a Listserv focused on chemical information has survived for 23 years. What is so special about chemical information that made all this happen?*

GW: I would have to say that it is not the chemical information per se that makes it special, but the chemists' desire to codify and make sense of their enormously complex scientific discipline. Chemical research sometimes leads to bewildering results and often to a mountain of data that needs to be interpreted and compared to previous results. Even in the early days of modern chemistry in the nineteenth century, chemists were devising ways to communicate and archive their results for future retrieval. Although other scientific disciplines have made similar efforts, it has long been recognized that chemists led the way in organizing their literature. I have sometimes used the analogy of Mendeleev's achievement in devising the periodic table to explain the impetus for cheminformatics. When faced with a large amount of data that was confounding, Mendeleev found a way to organize and mine it to give a logical and useful tool that even made predictions about elements that hadn't yet been discovered. Likewise today, with the data deluge that modern chemistry produces, the task of cheminformaticians and other chemistry knowledge workers is to make sense out of data and organize it in ways that allow future retrieval and analysis.

SB: *Is the way chemists perform research, seek information, and report their findings different from what researchers in other disciplines do?*

GW: By and large, the answer to this question is no. Chemists present preliminary results at conferences and publish the final results in scientific journals, as do scientists in other disciplines. However, the web has allowed scientists to bypass the traditional sequence of publishing scientific results, most noticeably in physics, where there is a tradition of posting preprints of papers before formal publication in a refereed scientific journal. A preprint archive in chemistry was tried, but it quickly failed because there was no tradition of exchanging preprints in chemistry. There is one aspect of chemistry that makes the information-seeking activities of chemists quite unique. That is the use of the chemical structure as a universally recognized search key. Much of the organization of chemical information is predicated on the similarities of properties among chemical substances with similar structures. It was only natural that ways

would be devised to search the primary chemical literature and secondary databases by structure. The culture of a scientific discipline is very hard to change. Nevertheless, when there is an overriding need that must be met, such as the need to clearly depict and search by a chemical structure, new developments will be adopted quickly by chemists. Thus, it can be predicted that the InChI (IUPAC International Chemical Identifier) will be accepted by chemists because it offers a nonproprietary identifier for chemical substances that can be used in printed and electronic data sources, allowing the linking of diverse data sources through the InChI.

SB: *Indiana University has a group working on electronic laboratory notebooks (ELNs) and their implementation in industrial and academic labs. As we know, the adoption of this technology has been slow in academic institutions. What obstacles exist (technological and psychological) for the acceptance of ELNs by researchers?*

GW: When I first came to Indiana University, I was responsible for a unit that produced microfiches of laboratory notebooks. It was never an idea that caught on in the Chemistry Department, but it did indicate that there was some consideration to getting the data into a format that was more easily portable and could be reproduced easily. One of our current PhD students, who is also employed in industry, has taken on some research in electronic laboratory notebooks. ELNs are an easier sell in industry because so much of the research in their labs is interrelated and driven by common goals of the company. Once a package is chosen, everyone uses it. In the academic setting, there are many research fiefdoms, and the research conducted in one group might have little or no relevance for other groups at that university. Hence, using the argument that ELNs must be used “for the good of the cause” is not very compelling in academia. On the other hand, all academic research groups have strong contacts with colleagues doing research at other academic institutions and national laboratories. It seems to me that there should be some incentive for closely related groups at multiple institutions to settle on a standard commercial ELN package that they would support. After all, it is the usual case that we see a lot of migration among such institutions by undergraduates who go on to graduate school, graduate students who accept postdoctoral fellowships at the related research groups, and faculty who move to new positions. It would make sense that one of the skills they take along with them is proficiency in the ELN software package of choice.

SB: *With research becoming more and more interdisciplinary and data-intensive, how do you see the ability of academic libraries, with their limited resources, to support eScience/eResearch in their institutions? Are the libraries promising something that they won't be able to deliver? And if someone needs to provide support for eScience, would the libraries be the suitable candidate to play this role?*

GW: Someone must support eScience/eResearch in academic institutions. However, a certain naivety about the data deluge can delude higher administrators in academic libraries into thinking that this is just another indexing and cataloging job, and we have been doing that for centuries. Designing a system for the effective retrieval and use of Big Data requires a tremendous understanding of the science and the likely uses to which the data will be put. With cloud computing becoming the norm in some areas of academic research, it is essential to understand the architecture of the current computing environment and to design effective and secure interfaces to those systems. Having said all of that, you might assume that I think academic librarians are

not up to the task of supporting eScience/eResearch, but that is not the case. In fact, I think that their traditions of service to scientific faculty are valuable foundations on which to build systems that are capable of dealing with the problem. However, it is not something that can be grafted onto the existing duties of librarians by appointing an associate dean for eScience/eResearch and charging that person to layer even more work onto already overworked librarians. If such efforts are going to be based in academic libraries, the administrators—both library and university administrators—must understand that the task will require much additional training and much more staff.

SB: *There are many discussions about the future of academic libraries, in general, and the new role that subject/liaison librarians will play in particular. Such new roles depend on reskilling of existing library staff and preparing a new generation of librarians trained to support research. In your opinion, are library schools in the country prepared to train librarians for such new roles?*

GW: Having worked as a professional academic librarian for nearly 35 years, I have thought a lot about whether the old skill sets still have value. What I have observed is that there is always an assumption by library deans that new duties will be added, but there is rarely a corresponding assumption that some of the old responsibilities will be taken away. To the extent that library schools have a good grasp of the technological advances being made to assist researchers and give the students at least an awareness of the full range of possibilities, including the most valuable of the old skill sets, they can provide an excellent training ground for new librarians. With MOOC [massive open online course] possibilities nowadays, they can even serve the retraining needs for librarians who are at various stages of their careers. The trend toward iSchools has been unrelenting in the last several decades. Indiana University was one of the last remaining bastions where a library school maintained a separate identity from the more computer-intensive informatics/computing science departments or schools. Now, the School of Library and Information Science at IU has been absorbed by the School of Informatics and Computing. I have long supported such a union of the two schools, and I anticipate a synergy that will lead to better preparation of practitioners along the whole spectrum from computer scientists and informaticians to information scientists and librarians.

SB: *Information literacy is a central focus for academic librarians. With the technological advancements and the dramatic changes in users' information needs and expectations, what role (if any) could librarians play in this new learning environment and what obstacles might they encounter in trying to adjust to these new developments?*

GW: I was glad to see information literacy come into its own as a defined specialization in academic libraries in recent years. Whatever technological changes may occur in the future, a fundamental approach to information acquisition and evaluation that goes far beyond googling should always be taught. My old mentor Herb White used to say about scientists doing their own online searches, "Yes, they do it, but they do it badly!" Today's end users have bypassed the command-driven approaches to searching online databases and are familiar only with user-friendly front-end systems such as SciFinder. Nevertheless, there are many intricacies of the Chemical Abstracts and MEDLINE databases that are masked by such a system. The single most important thing that people who are charged with teaching information literacy can do is to make budding scientists aware of the rich range of information tools available to them and

open their eyes to the advantages of choosing those tools over Internet search engines for almost all questions or information needs they might have. Certainly, librarians are most well suited to play the key role in information literacy on campus. The main obstacle they face is the faculty members who believe that the approach to information they have always used is the one that they should transmit to their students. It takes a gentle sales job to convince faculty members to try another approach and if not delegate the information literacy instruction to librarians, then at least share it with them. Although I have been out of the profession for some years now, it is safe to conjecture that there are still economic barriers to providing adequate information literacy training in some academic institutions. Librarians are best suited to negotiate with database publishers reasonable fees for the educational use of their products.

SB: *Some of the people who started the Open Access Movement—Vitek Tracz, for example—have moved on to other ventures such as Faculty of 1000 (<http://f1000.com>) and F1000Research (<http://f1000research.com>), which try to address issues in the “post-open access” world. There are predictions that the journal will disappear and be replaced by individual articles. In your opinion, how will the STEM publishing field change in the next few years?*

GW: I have always viewed the Faculty of 1000 as a special type of review serial. Any librarian who has ever had to conduct a serials cancellation project knows how fiercely the faculty will fight for the retention of review serials. I sometimes found that it was easier to cancel even an American Chemical Society journal than to remove an *Annual Review of* [fill in the blank] from our subscription list. The volume of primary scientific journal articles is immense, and scientists must have a mechanism that lets them focus on the articles most worthy of their attention. For years, they have depended on the reputation of a journal publisher, the editorial board of the journal, and the thoroughness with which the articles are subjected to peer review as key filters in their selection of the literature. Various alerting services have been developed to make sure they do not overlook a critical article, but even with those, key articles are sometimes missed. If we are to eliminate the journal as the unit of publication, then there must be something that substitutes for the confidence that a scientist has in the reputation of the journal itself. The move to electronic journals has opened the door to consideration of a system that would jettison the journal itself in favor of the article as the unit of publication. That may happen, but I do not foresee it coming to pass until there is a different architecture in place that gives the individual scientist a large measure of assurance that what is being read is not in fact junk science. The Faculty of 1000 is a novel way of assessing scientific journal output, but ultimately, there will be in place a mechanism that will assess the data itself behind a scientific article and give a rating to the scientist about both the validity of the research paper and the relevance of the research itself to a scientist's own research. There was at one time an attempt to characterize compilations of numeric data as being recommended (of the highest quality), provisional, typical, or selected. Data were assigned to these categories on the basis of checks against theoretical values, experimental detail, etc. What should ultimately emerge is a system using the power of cloud computing that allows such evaluation of data and articles to take place automatically and to be tied to an alerting service when items of potential interest are identified.

Publications by Gary Wiggins:

- Lee, W.M., Wiggins, G., 1997. Alternative methods for teaching chemical information to undergraduates. *Sci. Technol. Libr.* 16(3/4), 31–43.
- Wiggins, G., 1995. CHMINF-L, the chemical information sources discussion list. *J. Am. Soc. Inf. Sci.* 46(8), 614–17.
- Wiggins, G., 1998. Chemistry on the Internet: the library on your computer. *J. Chem. Inf. Comput. Sci.* 38(6), 956–65.
- Wild, D.J., Wiggins, G.D., 2006. Challenges for chemoinformatics education in drug discovery. *Drug Discov. Today* 11(9–10), 436–9.
- Wild, D.J., Wiggins, G.D., 2006. Videoconferencing and other distance education techniques in chemoinformatics teaching and research at Indiana University. *J. Chem. Inf. Model.* 46(2), 495–502.